MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963

(12)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TR38		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Tensorial Calibration. II. Second Order Tensorial Calibration	Technical Report - Interim
	6. PERFORMING ORG. REPORT NUMBER

AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Eugenio Sanchez and Bruce R. Kowalski	N00014-75-C-0536

PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Laboratory for Chemometrics Department of Chemistry BG-10 University of Washington	NR 051-565

1. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Materials Sciences Division Office of Naval Research Arlington, Virginia 22217	October 12, 1987
	13. NUMBER OF PAGES
	29

4. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

DTIC
SELECTED
OCT 26 1987
D
⊔D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

calibration	PCR	GRAM
tensor	MLR	
multivariate	PLS (partial least squares)	
second order	multidimensional arrays	
regression	generalized rank annihilation	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Tensorial calibration provides a useful approach to calibration in general. For calibration of instruments that produce two-dimensional (second order) data arrays of data per sample, tensorial concepts are a natural way of solving the calibration problems as vectorial concepts are for multivariate problems. Similarly, for third- and higher-order data, the tensorial description of calibration is also useful.

(over)

87 10   12

AD-A186 614

This paper introduces second order calibration from a tensorial point of view. Univariate, multivariate and bilinear approaches to calibration are presented. The generalized rank annihilation method (GRAM) is described from the tensorial perspective, and it is shown that GRAM is equivalent to finding a second order tensorial base that spans both tensors (calibration and unknown) with respective diagonal component matrices. GRAM uses a single calibration sample for multicomponent analysis even in the presence of interferences. Second order bilinear calibration is extended to multiple calibration samples, where the effect of collinearities is reduced.

Tensorial Calibration.

II. Second Order Tensorial Calibration

by

Eugenio Sanchez and Bruce R. Kowalski

Prepared for Publication

in

Journal of Chemometrics

University of Washington
Department of Chemistry BG-10
Seattle, Washington 98195

October 12, 1987

DTIC
COPY
INSPECTED
7

Accesion For

NTIS   CRA&I
DTIC   TAB
Unannounced
Justification

By
Distribution /

Availability Codes

Dist   Avail and / or
       Special

A-1

TENSORIAL CALIBRATION.
II. SECOND ORDER CALIBRATION

EUGENIO SANCHEZ AND BRUCE R. KOWALSKI

*Laboratory for Chemometrics, Department of Chemistry BG-10, University of Washington
Seattle, Washington 98195, U.S.A.*

SUMMARY

Tensorial calibration provides an useful approach to calibration in general. For calibration of instruments that produce two-dimensional (second order) data arrays of data per sample, tensorial concepts are a natural way of solving the calibration problems as vectorial concepts are for multivariate problems. Similarly, for third- and higher-order data, the tensorial description of calibration is also useful.

This paper introduces second order calibration from a tensorial point of view. Univariate, multivariate and bilinear approaches to calibration are presented. The generalized rank annihilation method (GRAM) is described from the tensorial perspective, and it is shown that GRAM is equivalent to finding a second order tensorial base that spans both tensors (calibration and unknown) with respective diagonal component matrices. GRAM uses a *single* calibration sample for multicomponent analysis even in the presence of interferences. Second order bilinear calibration is extended to *multiple* calibration samples, where the effect of collinearities is reduced.

INTRODUCTION

The complexity and amount of data generated by modern instruments for chemical analysis is in continuous increase. In particular, instruments that generate two-dimensional arrays of data (second order instruments) are now commonplace in the analytical laboratory. Time decay and emission-excitation fluorescence, chromatography-spectroscopy combinations, MS-MS and 2D-NMR, are a few of the many "hyphenated methods"[1,2] that generate such data. These instruments have become very important for the analyst mainly because their higher selectivity and resolution of signals, allowing for analysis of mixtures. Most of the time it is necessary to analyze several samples, either calibration or unknown samples. This additional complication implies that analysis of data from second order instruments may involve handling a three-dimensional (third order) array of data.

The normal way to handle this kind of data has been to choose from the array a single element that is unique for the analyte of interest, discarding or not collecting the rest of the data. For example, in MS-MS, is often possible to find daughter spectra ions that are completely unique for one analyte of a mixture. For an emission-excitation matrix (EEM), it is sometimes possible to find a combination of excitation and emission wavelengths for which only the analyte of interest has a significant signal.

With the increased popularity of multichannel (first order) instruments such us FT-IR spectrometers, the use of multivariate analysis in chemistry has increased in the last few years. Research in multivariate calibration[3] of near infrared reflectance data[4,5] and curve resolution[6,7] in chromatography are examples of this fact. The main advantage of using these kind of instruments is that it is not necessary to have a unique signal to determine an

analyte, by using the tools of multivariate calibration, e.g., multiple linear regression. Second order instruments have also benefit from multivariate analysis. In the simplest case, a second order instrument produces a matrix of data that can be seen as a series of first order data vectors. In fact, most second order instruments are a combination of two first order instruments. E.g., in LC/UV, a UV spectrometer (first order) is combined with a chromatograph (first order) to produce the instrument. Each scan can be seen as an individual UV spectrum, and at each wavelength an individual chromatogram is obtained. Therefore there are two possible multivariate (vectorial) spaces simultaneously: the spectral and the chromatographic spaces. The matrix of data can be equivalently represented in one or the other as a set of vectors (or points) in a multidimensional space. Fig 1 illustrates this point with an example.

A feasible approach for using multivariate calibration for second order instruments is to unfold the matrix of data in a long vector. In this way, the calibration matrix is a set of long vectors, and all the standard equations for multivariate calibration can be applied. Wold and coworkers have recently presented this approach, as an extension of the ordinary principal component (PC) and partial least squares (PLS) decompositions to sets of data from second and higher order instrumentation.[8]

Even with all the value of the multivariate approach to second order instruments, a serious limitation arises when using a vectorial representation for them. Multivariate calibration is not useful for determining the concentration of analytes in samples with unknown interfering constituents that were not accounted for in the calibration step. But Ho, Christian and Davidson have shown that for certain second order instruments this is not a problem.[9,10]

Conceptually, using a vectorial representation to describe second order data is similar to using univariate relations for multivariate data, by taking multichannel information

(vectorial) as a series of univariate data. In this case, it would reduce multichannel instruments to the limitations of univariate (single channel) instruments.

Just like vectors are a an extension from scalars, there is an extension from vectors that will describe second and higher order data, and that extension is *tensors*. Tensors encompass scalars (zero order tensors), vectors (first order tensors) and any higher orders (second- or higher order tensors) of data matrices. The first paper in these series was an introduction to a tensorial approach to calibration using well-known first order (multivariate) calibration[11]. This paper will use some of those ideas and concepts to develop second order calibration, and show h w for certain kinds of data it has considerable advantages over first order calib .

NOMENC .TURE

Boldface, capital letters represent tensors and matrices, e.g., $\mathbf{A}$. For a given matrix $\mathbf{A}$, the matrices $\mathbf{A}^T$, $\mathbf{A}^{-1}$, $\mathbf{A}^-$, and $\mathbf{A}^+$ stand for its transpose, inverse, generalized inverse and Moore-Penrose pseudoinverse, respectively. In general, an alternative representation for any tensor is given by its tensor components, italized, e.g., x_i, X_{ij}, M_{ijk}, $M_{ij..n}$, for a first, second, third and n^{th} order tensor respectively, either subscript or superscript. A first order tensor is equivalent to a vector. Boldface, lowercase letters represent first order tensors and column vectors, $\mathbf{v}$, and their transpose for a row vector, $\mathbf{v}^T$. A matrix, e.g., $\mathbf{V}$, and the set of its column vectors $\{\mathbf{v}_i\}$ are conveniently designated with the same letter, in upper case and indexed lowercase, respectively.

A standard summation convention is used to simplify nomenclature. When an index is repeated more than once only in one side of an equation, it implies a summation over the index valid range.[12]

To avoid confusion of terms, certain definitions have to be established as they are used

in this paper:

Order refers to the different directions or ways in a multi-order array or tensor. Examples of different orders of tensors and instruments were given in the last section. Synonyms of order used in the literature include "ways", "modes" and "spaces". We may refer to, e.g., the spectral order as well as the spectral space.

Dimensionality is the number of elements in a given order. For example, if one order corresponds to the UV spectrum, then the dimensionality of that order is the number of different wavelengths that are measured. For a tensor that has several orders, each order may have a different dimensionality.

Constituents are the chemical components in a sample that produce a response to the instrument; *Analytes* are those constituents that are being analyzed.

Components refers to the elements of a tensor in a given set of bases. For example, the elements of a vector are its components. Two kinds of components are defined: covariant and contravariant.[11,12] Also, in connection with principal components analysis, component refers to the principal component vectors.

The *rank* in a given order of a tensor will be defined as the number of varying independent factors in that order. The rank of an order cannot be greater than the dimensionality in that order.

The rank of a tensor will be defined as the number of varying independent factors in that tensor, above the noise level. For a non-null first order tensor (vector), the rank is defined as one; for a second-order tensor, the rank is equivalent to the rank of its matrix of components; For third and higher order tensor, the rank can be higher than the maximum dimensionality.[13]

LINEAR SECOND ORDER TENSORIAL CALIBRATION MODEL

From tensorial theory, it is known that the components of a second order tensor are also a matrix. Therefore, data acquired from a second order instrument can be defined as the components of a tensor. Designating M_{ij} the element at the i row and j column of a the data matrix, then there is an associated tensor $\mathbf{M}$ whose components are M_{ij}:

$$\mathbf{M} = \sum_{ij} M_{ij} \ \mathbf{a}_i \ \mathbf{b}_j \tag{1}$$

where $\mathbf{a}_i$ represents the basis vectors for one order and $\mathbf{b}_j$ represents the set of base vectors in the other order. They are the standard (physical) bases; e.g., for a LC/UV instrument, $\mathbf{a}_i$ represents measurement at time i, $\mathbf{b}_j$ represents measurement at wavelength j, and, M_{ij} is the instrument response at time i of the jth wavelength. Using the summation convention, Eq 1 is equivalent to

$$\mathbf{M} = M_{ij} \ \mathbf{a}_i \ \mathbf{b}_j \tag{2}$$

The summation convention invariably applies to summation over a repeated index on one side of the equation.

The response matrix $\mathbf{M}$ may be a function of many factors, and the concentrations of the analytes present in the sample are the most important one. Assuming that the response of a mixture is equal to the sum of the responses from the individual constituents, and there is a direct linear relationship between the responses and the concentration of the analytes, then $\mathbf{M}$ can be modeled as

$$\mathbf{M} = \sum_{i=1}^{q} c_i \mathbf{N}_i \ + \ \mathbf{N}_b \ + \ \mathbf{E} \tag{3}$$

where $\mathbf{M}$ is the response matrix of a second order instrument, to a sample with q analytes, c_i is the concentration of analyte i, $\mathbf{N}_i$ is the instrument matrix of responses for a pure analyte i; $\mathbf{N}_b$ is the background signal, and $\mathbf{E}$ is the model error. The background term can be either subtracted initially or inserted in the summation by defining it as the response

of an imaginary analyte with unitary concentration,

$$M = \sum_{i=1}^{q} c_i N_i + E \tag{4}$$

BASES AND COMPONENTS

The physical base, $\{a_i\}$ and $\{b_j\}$, is not the only possible base for describing the data. The data matrix M_{ij} from a second order instrument has an associated tensor M (see Eq 2). From tensorial theory, the tensorial object M is invariant to changes in base, therefore any base that spans the spaces defined by the standard bases a_i and b_j could be used to represent the tensor M, and it would be describing the same mathematical object. Two sets of bases useful for calibration will be introduced next.

Mathematical Base. A generally useful base for a second order tensor can be obtained from the singular value decomposition[14] (SVD) of its matrix of components. The SVD decomposes any matrix into the product of three matrices,

$$M = U S V^T, \tag{5}$$

where U and V are orthonormal matrices, i.e., their respective columns $\{u_i\}$ and $\{v_j\}$ are orthogonal, unitary vectors; and S is a diagonal matrix. These two sets of vectors can be used as the base for the tensor M, and the covariant and contravariant components are equivalent, because they are orthonormal, and a direct product (projection formula) is enough for the change of base. In matrix notation, by left multiplying M by U^T and right multiplying by V, the components of M in the new base are obtained.

$$M_{uv} = U^T M V = U^T U S V^T V = S \tag{6}$$

But S is a diagonal matrix, therefore the singular value decomposition vectors provide an orthonormal base where the components of the tensor are a diagonal matrix, and Eq 2 is reduced to

$$M = S_i \, u_i \, v_i \tag{7}$$

where the summation runs over the i index. Thus, the matrices M and S represent the same tensor under different bases. There is an infinite number of matrices that represent the tensor M, but in most cases there is only one that is, at the same time, diagonal and its bases are orthonormal, and it is given by the singular value decomposition. Fig 2 illustrates the idea of differents bases for the same second order tensor.

Chemical Base. There usually are relationships within the data matrix that further constrain the possible values of the data. The fact that the elements of a column are related, e.g., because they where collected simultaneously, is just one of those possible relationships. A very important case occurs for data that is bilinear.[15] Such data has the interesting property that if there is only *one* analyte present in the sample, then the matrix can be approximately factorized as the outer product of two vectors. This also implies that the rank of the matrix should be one, within the noise level. If the technique is, e.g., emission-excitation fluorescence, then one of the vectors would correspond to the emission spectrum of the analyte and the other to its excitation spectrum,

$$N_i = x_i y_i^T + E_i \tag{8}$$

where N_i is the bilinear data matrix for an analyte at unitary concentration, x_i is the spectrum in one order, e.g., excitation; y_i is the spectrum in the other order, e.g., emission; and E_i is the error matrix of the approximation. This base is of great interest to the chemist dealing with bilinear data. If a sample has q analytes, then the matrix can be modeled by a sum of q unitary rank matrices, and will have rank q,

$$M = \sum_{i=1}^{q} c_i N_i = \sum_{i=1}^{q} c_i x_i y_i^T \tag{9}$$

where the error in the model has been dropped for simplicity. Rewriting Eq 9 in matrix notation, by considering the vectors $\{x_i\}$ and $\{y_i\}$ as the columns of the matrices X and Y, and defining C as a diagonal matrix, $\text{diag}(C) = \{c_1, c_2, \ldots, c_q\}$, yields

$$M = X C Y^T + E \tag{10}$$

scaling the $\{x_i\}$ vectors and the $\{y_i\}$ vectors to be of unitary length, then the matrix C has to be scaled to another matrix, with the normalization constants in its diagonal, and zeros in the rest (for simplicity, the new normalized matrices will also be denominated X and Y),

$$M = X \beta Y^T \tag{11}$$

This equation is similar to Eq 5, with a diagonal matrix ß, but the matrices X and Y are not orthonormal. They are the pure analyte spectra, or sensor array patterns, in each order. By analogous reasoning, expressing the tensor M in the bases $\{x_i\}$ and $\{y_i\}$, its components will be again a diagonal matrix, in this case ß. It is important to realize that Eq 11 is only an approximation. For normal experimental conditions, factors such as noise, or systematic deviations from the model are always present. Similarly to Eq 9, the tensor M can be approximated as

$$M \approx \beta^{ii} x_i y_i = \beta_i x_i y_i \tag{12}$$

Note the position of the indexes ii as a superscript, indicating that the β^{ii} are contravariant components for both bases. If the instrument is bilinear, the vectors x_i and y_i do not change with a change in concentration of the analyte i, i.e., c_i. The only thing that changes is β^{ii}, therefore, for a linear response instrument, the β^{ii}'s are directly proportional to the concentrations, and if they are determined, they can be used for quantitation.

CALIBRATION OF SECOND ORDER INSTRUMENTS

Univariate Calibration

Quantitation with second order instruments is often performed by selecting one response from the matrix that is completely unique the the analyte of interest. That response is only a function of the concentration of that analyte, if Eq 9 holds, the response

is linearly related to the concentration

$$. M_{ij} = a\, c_k + b \tag{13}$$

The advantage of this method is its simplicity, in many cases requires only to collect the specific signal that is unique, rather than all the matrix of data. The disadvantage is that it has the same limitations that other univariate techniques have, i.e., it is impossible to detect or correct for interferences.

Multivariate Calibration

It is possible to use multivariate calibration for building a concentration prediction model. Several or all the responses from the matrix can be arranged as a vector of responses,

$$i,j = i\,(l), j\,(l) \qquad\qquad r_l = M_{ij} \tag{14}$$

and the calibration problem can be reduced to the multivariate case, with every sample having a response vector **r**. Geladi and coworkers[8] have use this approach for the second and higher order data, unfolding all the data matrix into the vector r. Multivariate calibration is well understood, and a tensorial approach to it has also been discussed in the previous paper of these series,[11] therefore will not be emphasized here.

There are several advantages in reducing second order calibration to a vectorial, multivariate problem. The analyst may choose a few relevant responses from the matrix, e.g., the first row, reducing the acquisition time. It will have the advantages of multivariate calibration, e.g., correction for background present in the calibration samples and detection of interferences. In addition, this method can be used with data from any second order instrument, whereas the methods discussed latter in this section are specific for bilinear data. For non-bilinear data, multivariate calibration is the only choice at the present time.

BILINEAR CALIBRATION

It is natural to suspect that if going from zero order to first order calibration there are advantages, then there might be advantages for going from first order to second order instruments. As it turns out, for bilinear second order data this is true, and it will be described in this section.

The problem of calibration of second order bilinear data can be presented as one of finding the components of the response matrix on two specific sets of base vectors, corresponding to the responses in each order from a unitary amount of the analytes present in the sample.

For example, if there are only two analytes present in a sample, the response matrix M can be modeled as a simple linear combination of the individual analytes responses in each order,

$$M = c_1 \; x_1 \, y_1^T + c_2 \; x_2 \, y_2^T \tag{15}$$

if the matrices $x_k y_k^T$ were defined as the pure analyte responses at unitary concentration of analyte i, then the tensorial components of M in the *bases* $\{(x_1, x_2); (y_1^T, y_2^T)\}$ would simply be the concentrations c_1 and c_2.

In a similar way to calibration for first order instruments,[11] bilinear calibration involves finding the components of the response matrix on two specific sets of base vectors. To use a projection formula for finding the concentrations of analyte i, two contravariant vectors must be estimated: x_k^* and y_k^*,

$$c_k = x_k^{*T} M y_k^* \tag{16}$$

(No summation applies because index k is in both sides of the equation). These contravariant vectors are unknown, but it is possible to estimate c_i from a known tensor that is closely related to M: its pseudoinverse M^+,

$$M^+ = (XCY^T)^+ = (Y^T)^+ C^{-1} X^+ \tag{17}$$

$$(M^+)^T = (1/c_k) x_k^* y_k^{*T} \tag{18}$$

Therefore, the pseudoinverse of M is a linear combination of the contravariant vectors for the analytes present in the sample, and the coefficients are the diagonal matrix of the concentration inverses. To estimate this coefficients, we need the covariant vectors corresponding to the analyte of interest, i.e. their pure spectra x_k and y_k at unitary concentration

$$1/c_k = y_k^T M^+ x_k \tag{19}$$

Single Sample Bilinear Calibration

If we measure a single calibration response matrix for a pure analyte, $N_k = c_{N,k} x_k y_k^T$, the concentration ratio can be estimated as

$$c_{N,k}/c_{M,k} = c_{N,k} y_k^T M^+ x_k = N_{ij} (M^+)_{ij} \tag{20}$$

where a double summation over i, j applies, and $(M^+)_{ij}$ represents the ith row and jth column component of the pseudoinverse of M. Note that this equation does not include any calibration information about the other analytes present in the sample, i.e., quantitation is possible in the presence of unknown interferences.

Simultaneous Estimation of Concentrations and Spectra.

If the calibration sample contains several analytes, Eq 20 will yield erroneous results. But it is possible to show that quantitative analysis is still feasible for this case, by taking a different approach.

Assume that there is only one *multicomponent* calibration sample. Calling N the tensor of its responses, it can be modeled with an equation similar to Eq 11,

$$N = X \xi Y^T + E_N \tag{21}$$

N is the tensor in the standard base and ξ is its corresponding components in the X, Y

bases. The components of N are known only in the standard base. To describe a general case, assume that the X and the Y are matrices that contain the superset of all the components present in *both* M and N. There will be corresponding β_i or ξ_i elements that will be zero if they are not present in one or the other sample. It is evident that the bases X and Y can approximate both tensors M and N with a diagonal matrix of components, respectively ß and ξ. It will be shown that under certain conditions, there is only one base that diagonalizes both tensors simultaneously, and that base can be estimated, together with the ratios of the diagonal elements ξ_i/β_i.

To find the base, Eqs 10 - 11 can be used as a system of two matrix equations with four unknowns, namely X, Y, ß and ξ. A possible approach to solve this system is to express N as a function of M,

$$N = X \xi Y^T = X ß^{-1} \xi ß Y^T$$

$$N = X (ß^{-1} \xi) X^+ (X ß Y^T)$$

$$N = X (ß^{-1} \xi) X^+ M \tag{22}$$

these equations are valid only if all the elements of the diagonal matrix ß are non-zero because its inverse ($ß^{-1}$) must be computed. This implies that M contains all the components of the superset X and Y. For a case that this is not true, the new matrix W = M + N can be used instead of M, that by definition would include all the components. X^+ represents the contravariant form of the base X, and is simply the generalized inverse of the X matrix. Defining $\lambda \equiv ß^{-1} \xi$, Eq 13 can now be factorized to result into a non-symmetric eigenvalue-eigenvector problem, after right multiplying by M^+ and then X as,

$$N M^+ = X \lambda X^+ M M^+ \tag{23}$$

$$(N M^+) X = X \lambda X^+ X$$

$$(N M^+) X = X \lambda \tag{24}$$

The spectra X are the *right* eigenvectors of the square non-symmetric matrix ($N M^+$), and

the eigenvalues λ are the ratios of concentrations, because β and ξ are proportional to concentrations. Once X is known, Y^T can be estimated using

$$Y^T = \beta^{-1} X^+ M \qquad (25)$$

Eqs 24 and 25 summarize the Generalized Rank Annihilation Method (GRAM).[15]

A particular case of Eq 24 arises when N has only one component, $N = x_1 \xi_1 y_1^T$. All the eigenvalues will be nearly zero with the exception of one, $\lambda_1 = \beta_1/\xi_1$. Then Eq 24 can be rewritten as

$$(x_1 \xi_1 y_1{}^T M^+) x_1 = x_1 \lambda_1 \qquad (26)$$

dropping x_1 and changing sides,

$$\lambda_1 = \xi_1 y_1{}^T M^+ x_1 \qquad (27)$$

which is equivalent to the non-iterative Rank Annihilation equation introduced by Lorber.[16,17] Actually it is not necessary to estimate $y_1{}^T$ and x_1 from N. Recognizing that $N_{ij} = \xi_1 x_i y_j$, Eq 27 can be simplified to

$$\lambda_1 = N_{ij} (M^+)_{ij} \qquad (28)$$

which is equivalent to Eq 20.

Characteristics of Bilinear Calibration

Eqs 13-16 provide information useful to understand the possibilities and limitations of Bilinear Calibration using GRAM. First of all, they not only represent a calibration method, but also a curve resolution method, because the intrinsic factors, X and Y, are extracted. It is not the same curve resolution as described in the literature,[18,19] in which an uncertainty region is defined where the intrinsic factors are present, and further constraints must be used to choose a solution within the region. GRAM estimates a unique solution without empirical assumptions.

It is a fact that when two (or more) eigenvalues are identical (or very close to each other

for experimental data), their corresponding eigenvectors are not unique. Therefore, one of the restrictions of GRAM is that if two components have the same ratio of concentrations between the samples, their eigenvalues will be very similar, and the estimated spectra will be unreliable. Nevertheless, the eigenvalues should still provide quantitative information, but it will be more difficult to match the eigenvalue with its corresponding analyte, because no estimated spectrum is available. If a library spectrum is available, simple target factor analysis on the space generated by all the eigenvectors with the same eigenvalue should confirm the identity of at least one of the eigenvalues.[20,21]

A limitation arises when the intrinsic spectra in one of the orders are not linearly independent. Eq 17 is no longer valid, because $X \lambda X^+ MM^+$ is not equal to $X \lambda X^+$. Assuming that the Y are linearly dependent, the matrix M will have lower rank than the matrix X, which is made up of linearly independent vectors. The matrix (MM^+) is a projection matrix that behaves like the identity matrix for vectors in the subspace spanned by M. But if X has higher rank than M, it is necessarily true that the vectors in X will have some component outside of the space spanned by M, therefore MM^+ does not leave $X \lambda X^+$ unchanged. The results then should be unpredictable when such a dependency exists. Fortunately there is a simple test that will detect, but not correct this problem: the projection of N on $M M^+$ should leave N unchanged within the noise level.

The estimation of the pseudoinverse of M is the most important step in GRAM. In a similar way to first order tensorial calibration, the selection of the proper subspace for the pseudoinverse is a determining step in the quality of the GRAM results. Eq 18 shows that N is literally projected onto M, therefore the best way to span the space is to find a set of vectors that express both M and N in an unbiased way. The approach suggested in this work is to join the matrices in two different ways to form two larger matrix, and then obtain their singular value decomposition,

$$W_1 = (M|N) = U_1 \, S \, V_1^T \tag{29a}$$

$$W_2 = (\tfrac{M}{N}) = U_2 \, S \, V_2^T \tag{29b}$$

where W_1 has twice as many columns as M, and the same number of rows; and W_2 has twice as many rows as M, and the same number of columns. The columns of the matrix U_1 are an unbiased estimation of the column subspace of both M and N, and the columns vectors V_2 will similarly be an unbiased estimation of the row space of both M and N. The number of relevant singular values is first estimated by cross validation in order to truncate the SVD matrices U_1 and V_2.[22,23] Defining the truncated matrices as $U = \bar{U}_1$ and $V = \bar{V}_2$, new projected matrices are used for the calculations,

$$N = UU^T N VV^T \tag{30}$$

$$M^+ = V (U^T M V)^{-1} U^T \tag{31}$$

$$N M^+ = UU^T N VV^T V (U^T M V)^{-1} U^T$$

$$N M^+ = U (U^T N V) (U^T M V)^{-1} U^T$$

$$N M^+ = U N_{uv} M_{uv}^{-1} U^T \tag{32}$$

where N_{uv} and M_{uv} stand for the N and M tensors components in the U, V subspaces. Finally, The new matrix $N M^+$ is then substituted in Eq 24 for a GRAM calculation.

MULTIPLE SAMPLES CALIBRATION: TRILINEAR DECOMPOSITION

Even with the advantage that only one sample is necessary for doing multicomponent determinations, in many cases it is desirable to use more than one calibration sample, to

(1) Cover a wide dynamic range of concentrations for each analyte. Deviations from the linear model may produce big prediction errors.

(2) Ensure that not two analytes will have the same concentration ratio, precluding the extraction of the spectra.

(3) Reduce the effect of collinearities.

(4) Increase the precision of the predicted concentrations.

The problem of bilinear multiple sample calibration can be solved in several ways, and it is equivalent to a trilinear decomposition. In general there will be a set of L samples, either calibration or unknown samples, with matrices of response

$$\{ \mathbf{M}_1, \mathbf{M}_2, \dots \mathbf{M}_l, \dots \mathbf{M}_L \}$$

Following Eq 11 or 21, any bilinear response matrix l can be modeled as

$$\mathbf{M}_l = \mathbf{X}\, \beta_l\, \mathbf{Y}^\mathbf{T} + \mathbf{E}_l \tag{33}$$

which is also a third order tensor with components

$$M_{ijl} = X_{ik}\, Y_{jk}\, Z_{lk} + E_{ijl} \tag{34}$$

where $Z_{lk} = (\beta_l)_{kk}$ and E_{ijl} is the error of the model. The trilinear decomposition of M_{ijl} is the matrices X_{ik}, Y_{jk} and Z_{lk}. If some of the matrices $\mathbf{M}_l$ are unknowns and others are calibration samples, it is possible to use Z_{lk} to estimate the relative concentrations. The trilinear decomposition is unique in many cases, and it is usually solved by minimizing the sum of squares of the residuals

$$R_{SS} = \sum_{ijl} E_{ijl}^2 = \sum_{ijl} (M_{ijl} - \sum_{k=1}^{q} X_{ik}\, Y_{jk}\, Z_{lk})^2 \tag{35}$$

Appellof and Davidson found that for trilinear chemical data it was possible to obtain the intrinsic vectors uniquely, by using a minimization algorithm and at least as many slices (matrices of data) in the third order as components were present in the mixture.[24] In fact, GRAM can be seen as a particular case of decomposition of a trilinear matrix, with only two slices in the third order, the calibration and the test bilinear data respectively.

The problem of trilinear decomposition was studied in the early seventies by researchers in the area of psycometrics.[25] It was discovered that the number of factors, K, that could be extracted uniquely from third order data of trilinear nature is related to the rank

in each order by the formula:

$$K \leq (q_x + q_y + q_z - 2)/2 \qquad (36)$$

where q_x, q_y, and q_z are the ranks in each order as defined by Kruskal[26] and the formula sets an upper limit for the number of factors that can be extracted. For example, for $q_z = 2$ (two slices in the third order), and $q_x = q_y$, it is concluded that $K \leq q_x$. This means that the upper limit for the number of unique factors that can be extracted is equal to the rank of the first two orders, which is the result found for GRAM, where the number of extracted components has to be equal or less than the rank in each order (by definition, for bilinear data, the rank in the two orders is the same).

Unfortunately, the available algorithms for trilinear decomposition are based in iterative minimizations of residuals, e.g., Alternating Least Squares procedures. Convergence is not always achieved, and the more components that are present, the more difficult it is to find the right solutions. This is a fundamental difference with GRAM, for which the intrinsic vectors are eigenvectors, and no iterations are necessary. However, it is possible to use GRAM in several ways to solve trilinear decompositions for bilinear calibration with multiple samples.[27]

CONCLUSION

The application of GRAM is not restricted to calibration. Whenever two or more samples that have some constituents in common are available, GRAM can be applied. For example, solvent extraction could be used to generate two samples out of a single unknown, and with a proper selection of the solvent, the analytes will have different ratios of concentration. Then a bilinear instrument and GRAM can be used to extract the spectra of those analytes for identification.

The greatest potential for GRAM and in general second order methods is perhaps in

future second order instruments which are yet to be built. When using a bilinear instrument with GRAM, the analyst need not worry about interferences and/or contaminants in the analysis, and simultaneous determination of several components is possible with only one calibration sample. Until recently, the main goal in analytical chemistry has been to increase the resolution more and more, without considering data in other orders, simply because no advantage was seen in doing so. That trend has started to change, and everyday more second order instruments become available, as well as the mathematical tools to handle their data.

Finally, the problem of second order calibration presents numerous challenges, such as sample and sensor selection selection for optimal calibration design, calibration of non-bilinear second order data and error propagation. Work in some of these areas has already started by the authors, and will be presented in future publications.[28]

BIBLIOGRAPHY

1   T. Hirschfeld, *Anal. Chem.* **52**, 297A (1980).

2   T. Hirschfeld, *Science* **230**, 230 (1985).

3   L.S.Ramos, K.R. Beebe, W.P. Carey, E. Sanchez, B.C. Erickson, B. Wilson, L.E. Wangen and B.R. Kowalski, *Anal. Chem.* **58**, 294R (1986).

4   I.A. Cowe and J.W. McNicol, *Appl. Spectrosc.* **39**, 257 (1985).

5   P. Geladi, D. MacDougall and H. Martens, *Appl. Spectrosc.* **39**, 491 (1985).

6   O.S. Borgen and B.R. Kowalski, *Anal. Chim. Acta* **174**, 1 (1985).

7   H. Gampp, M. Maeder, C.J. Meyer and A.D. Zuberbuehler, *Anal. Chim. Acta* **193**, 287 (1987).

8   S. Wold, P. Geladi, K. Esbensen and J. Öhman, *J. Chemometrics* **1**, 41 (1987).

9   C-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.* **50**, 1108 (1978).

10   C-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.* **52**, 1071 (1980).

11   E. Sanchez and B.R. Kowalski, *J. Chemometrics*, Submitted (1987)

12   B. Budiansky, *Tensors*, In: Pearson, C. E., Ed. *Handbook of Applied Mathematics*, Van Nostrand Reinhold Company, New York, ch. 4, 179 (1974).

13   J.B. Kruskal, *Psychometrika* **41**, 281 (1976).

14   C.L Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, N.J. (1974).

15   E. Sanchez and B.R. Kowalski, *Anal. Chem.* **58**, 496 (1986).

16   A. Lorber, *Anal. Chim. Acta* **164**, 293 (1984).

17   A. Lorber, *Anal. Chem.* **57**, 2395 (1985).

18   W.H. Lawton and E.A. Sylvestre, *Technometrics* **13**, 617 (1971).

19   D.W. Osten and B.R. Kowalski, *Anal. Chem.* **56**, 991 (1984).

20   E.R. Malinowski and D.G. Howery, *Factor Analysis in Chemistry*, Wiley, New York (1980).

21   A. Lorber, *Anal. Chem.* **56**, 1004 (1984).

22   H.T. Eastment and W.J. Krzanowski, *Technometrics* **24**, 73 (1982).

23   S. Wold, *Technometrics* **20**, 397 (1978).

24   C.J. Appellof and E.R. Davidson, *Anal. Chem.* **53**, 2053 (1981).

25   R.A. Harshman, *UCLA Working Papers in Phonetics* **16**, 1 (1970).

26   J.B. Kruskal, *Psychometrika* **41**, 281 (1976).

27   E. Sanchez and B.R. Kowalski, in preparation.

ACKNOWLEDGEMENTS

FIGURE CAPTIONS

Fig 1. Equivalent representations of a simulated LC/UV data matrix with rank =2. (A)
Twelve points corresponding to the scans in the UV spectral space, (B) Fourteen points
corresponding to the wavelengths in the chromatographic space.

Fig 2. Tensorial change of base. Representation of equivalent tensors in different bases.
R is the tensor in the standard bases and R_{uv} is the tensor in the bases U, V. If the
singular value decomposition of $R = U\ S\ V^T$, R_{uv} is the diagonal matrix S.

$R_{uv} = S$

$R = USV^T$

U

V

R
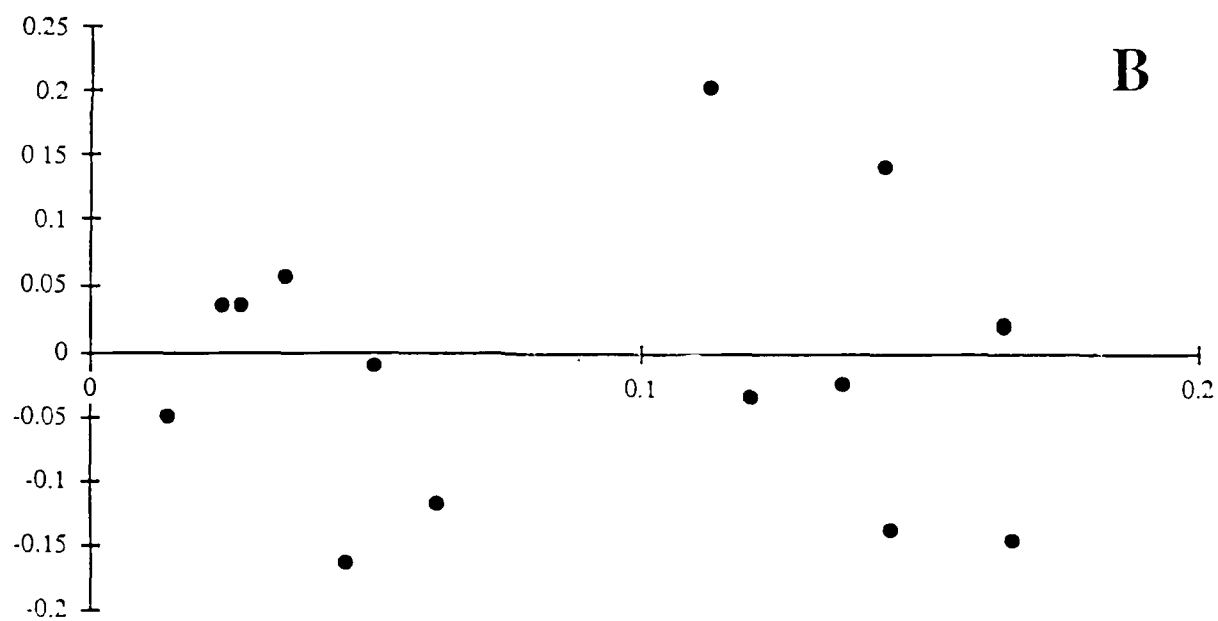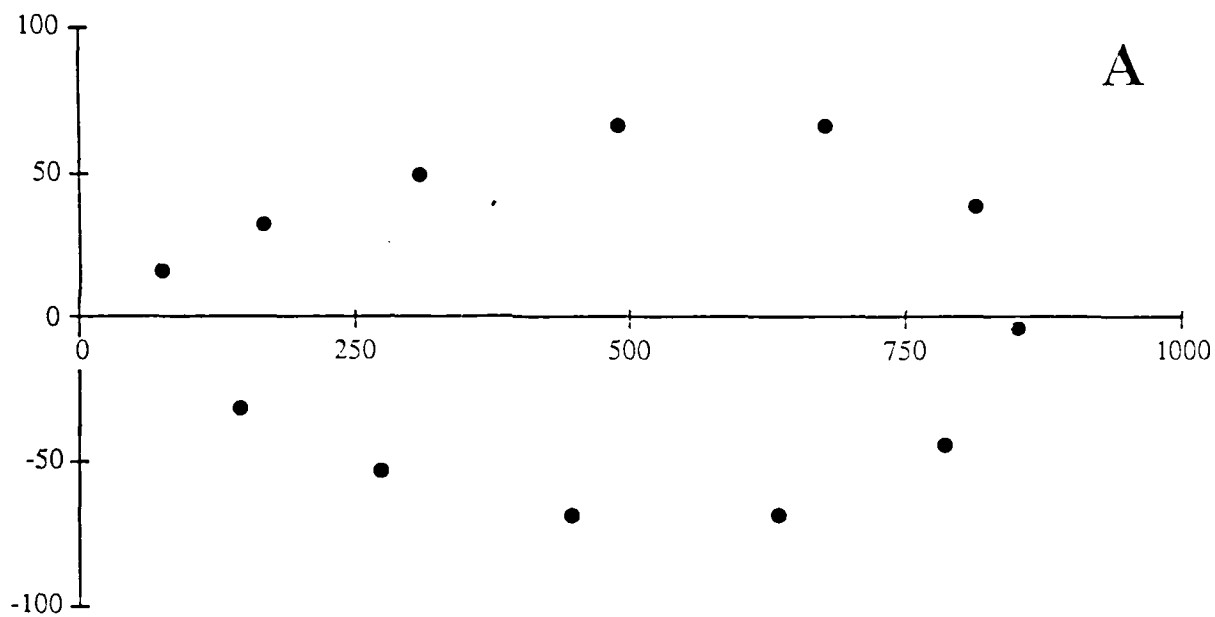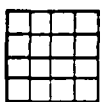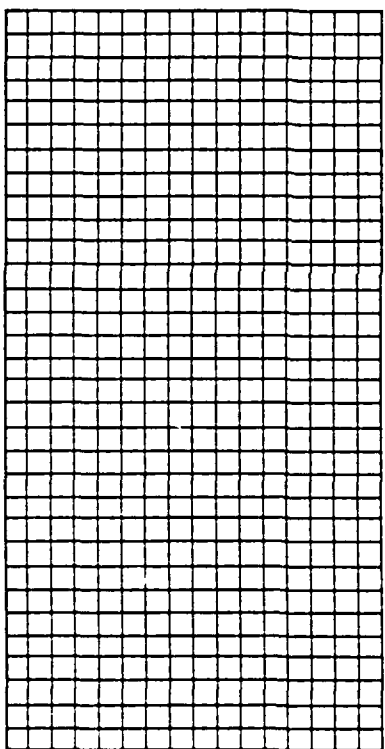
Office of Naval Research   2
Attn: Code 413
800 N. Quincy Street
Arlington, Virgina 22217

Dr. Bernard Douda
Naval Weapons Support Center
Code 5042
Crane, Indiana 47522

Naval Air Systems Command
H. Rosenwasser, Commander
Code 310C
Washington D.C. 20360

Naval Civil Engineering Lab
Attn: Dr. R.W. Drisko
Port Hueneme, CA 93401

Defense Technical Info. Ctr.   12
Building 5. Cameron Station
Alexandria, Virginia 22314

DTNSRDC
Attn: Dr. G. Bosmajian
Applied Chemistry Division
Annapolis, Maryland 21401

Dr. William Tolles
Superintendent
Chemistry Division, Code 6100
naval Research Laboratory
Washington D.C. 20375

Dr. David Young
Code 334
NORDA
NSTL, Mississippi 39529

Naval Weapons Center
Attn: Dr. A.B. Amster
Chemistry Division
China Lake, California 93555

Scientific Advisor Commandant
of the Marine Corps
Code RD-1
Washington D.C. 20380

U.S. Army Research Office
Attn: CRD-AA-IP
P.O. Box 12211
Research Triangle Park, NC
27709

Mr. John Boyle
Materials Branch
Naval Ship Engineering Center
Philadelphia, PENN 19112

Naval Ocean Systems Center
Attn: Dr. S. Yamamoto
Marine Sciences Division
San Diego. California 91232

Dr. R. A. Osteryoung
Department of Chemistry
State University of New York
Buffalo. New York 14214

Dr. J. Osteryoung
Department of Chemistry
State University of New York
Buffalo. New York 14214

Dr. H. Chernoff
Department of Mathematics
Massachusetts Institute of
Technology
Cambridge. MASS 02139

Dr. A. Zirino
Naval Undersea Center
San Diego. California 92132

Dr. George H. Morrison
Department of Chemistry
Cornell University
Ithaca. New York 14853

Dr. Alan Bewick
Department of Chemistry
Southampton University
Southampton. Hampshire
ENGLAND S09 5NH

Dr. M. B. Denton
Department of Chemistry
University of Arizona
Tucson. Arizona 85721

Dr. S. P. Perone
Lawrence Livermore Natl Lab
 L-370
P.O. Box 808
Livermore. California 94550

Dr. G. M. Hieftje
Department of Chemistry
Indiana University
Bloomington. Indiana 47401

Dr. Christie G. Enke
Department of Chemistry
Michigan State University
East Lansing. Michigan 48824

Walter G. Cox. Code 3632
Naval Underwater Systems Ctr.
Building 148
Newport. Rhode Island 02840

Professor Isiah M. Warner
Department of Chemistry
Emory University
Atlanta. Georgia 30322

Dr. Kent Eisentraut
Air Force Materials Lab
Wright-Patterson AFB. **OH**
45433

Dr. Adolph B. Amster
Chemistry Division
Naval Weapons Center
China Lake. California 93555

Dr. B. E. Douda
Chemical Sciences Branch
code 50 C
Naval Weapons Support Center
Crane. Indiana 47322

Dr. John Eyler
Department of Chemistry
University of Florida
Gainesville. Florida 32611

Professor J. Janata
Department of Bioengineering
University of Utah
Salt Lake City. Utah 84112

Dr. J. DeCorpo
NAVSEA
Code 05R14
Washington D.C. 20362

Dr. Charles Anderson
Analytical Chemistry Division
Athens Environmental Lab
College Station Road
Athens. Georgia 30613

Dr. Ron Flemming
B 108 Reactor
National Bureau of Standards
Washington D.C. 20234

Dr. Frank Herr
Office of Naval Research
Code 422CB
800 N. Quincy Street
Arlington, Virginia 22217

Professor E. Keating
Dept. of Mechanical Engineering
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. M. H. Miller
1133 Hampton Road
Route 4
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. Clifford Spiegelman
National Bureau of Standards
Room A337 Bldg. 101
Washington D.C. 20234

Dr. Denton Elliott
AFOSR/NC
Bolling AFB
Washington D.C. 20362

Dr. B. E. Spielvogel
Inorganic and Analytical Branch
P.O. Box 12211
Research Triangle Park, NC
27709

Ms. Ann DeWitt
Material Science Department
160 Fieldcrest Avenue
Raritan Center
Edison, New Jersey 08818

Dr. A. Harvey
Naval Research Laboratory
Code 6110
Washington D.C. 20375

Mr. S. M. Hurley, Code 032P
Naval Facil. Engineer. Command
200 Stovall Street
Alexandria, Virginia 22331

Ms. W. Parkhurst, Code R33
Naval Surface Weapons Center
Silver Spring, Maryland 20910

Dr. M. Robertson, Code 305
Electrochem Power Sources Div.
Naval Weapons Support Center
Crane, Indiana 47522

Dr. Andrew T. Zander P1204
Perkin-Elmer Corporation
901 Ethan Allen Hwy/ MS905
Ridgefield, Connecticut 06877

Dr. Marvin Wilkerson
Naval Weapons Support Center
Code 30511
Crane, Indiana 47522

Dr. J. Wyatt
Naval Research Laboratory
Code 6110
Washington D.C. 20375

Dr. J. MacDonald
Naval Research Laboratory
Code 6110
Washington D.C. 20375

Dr. H. Wohltjen
Naval Research Laboratory
Code 6170
Washington D.C. 20375

Dr. John Hoffsommer
Naval Surface Weapons Center
Building 30, Room 208
Silver Spring, Maryland 20910

Dr. Robert W. Shaw
U.S. Army Research Office
Box 12211
Research Triangle Park, NC
27709

END
DATE
FILMED

JAN
1988